

## Molecular and expression analysis of a family of the *Amblyomma americanum* tick Lospins

Albert Mulenga\*, Rabuesak Khumthong and Maria A. Blandon

Department of Entomology, College of Agriculture and Life Sciences, Texas A & M University, TAMU 2475, College Station, TX 77843, USA

\*Author for correspondence (e-mail: a-mulenga@tamu.edu)

Accepted 28 June 2007

### Summary

Serine proteinase inhibitors (serpins) are a family of structurally similar but functionally diverse proteins that regulate several important proteolytic cascades in most branches of life. We have characterized 17 *Amblyomma americanum* serpin cDNAs here named as 'Lospins' (L; an acronym for Lone Star tick serpin) that possess three  $\beta$ -sheets, eight  $\alpha$ -helices and a reactive center loop consistent with the consensus serpin superfamily secondary structures. Visual inspection of deduced amino acid sequences revealed two patterns of basic residues: (i) <sup>86</sup>DKSRVLKAYKRL<sup>97</sup> in L5 and L13–16 and (ii) <sup>158</sup>VRDKTRGKI<sup>166</sup> in all Lospins, which are similar to consensus glycosaminoglycan (GAG) binding sites (XBnXmBX, where X and B are non-basic and basic residues, n=1 or 2 and m=1, 2 or 3). On three-dimensional models, the two putative GAG binding sites mapped onto  $\alpha$ -helices D and F, respectively, with calculation of

electrostatic surface potentials revealing basic patches on L5 and L13–16 models that are comparable to the heparin-binding site on antithrombin. RT-PCR expression analysis of 15 selected genes showed that the majority (11/15) of the Lospins were ubiquitously expressed in the midgut, ovary and salivary glands. On a neighbor-joining phylogeny guide tree, 15 serpins from other ticks and 17 Lospins from this study, a total of 32 tick serpin sequences, segregated into five groups with Lospins in groups A and D being conserved across tick species. The discovery of Lospins in this study sets the framework for future studies to understand the role of serpins in tick physiology.

Key words: *Amblyomma americanum*, serine proteinase inhibitors (serpin), tick physiology glycosaminoglycan (GAG) binding site, tick vaccine.

### Introduction

Ticks are among the most important vectors of disease agents affecting both humans and animals, and are currently considered to be vectors of human infectious diseases in the world second only to mosquitoes (Sonenshine, 1993). Control of ticks has traditionally been accomplished by use of acaricides, which has resulted in selection of resistant ticks and environmental pollution. Of the various proposed alternatives to acaricide use (Sonenshine, 1993), vaccination has emerged as a sustainable, cost-effective and environmentally friendly alternative. A deeper understanding of tick molecular physiology is needed before development of alternative tick control methods can be achieved. The current focus of tick research is to uncover the molecular basis of tick physiology (Hill and Wikel, 2005; Nene et al., 2002a; Nene et al., 2002b; Ribeiro et al., 2006).

One group of proteins that may play important roles in tick physiology are serine proteinase inhibitors (serpins). Serpins represent one of the largest superfamilies of proteins found in most branches of life, ranging from viruses to vertebrates (Huntington, 2006). In humans, the majority of serpins function as negative regulators of several tightly regulated pathways such as blood coagulation, inflammation, complement activation, cancer metastasis and food digestion (Silverman et

al., 2001; Huntington, 2006). More than 90 human diseases result from natural mutations of serpins, which attests to the importance of this family of proteins in the physiology of multicellular organisms (Potempa et al., 1994; Silverman et al., 2001). Although it cannot be assumed that observations in mammalian serpins will also be true for ticks, we are encouraged by evidence from other invertebrate systems where serpins have been linked to regulation of important pathways such as innate immunity (Abraham et al., 2005; Michel et al., 2005; Pelte et al., 2006; Nappi et al., 2005; Zou and Jiang, 2006) and embryo development (Carrell and Corra, 2004; Rushlow, 2004).

Given the importance of serpins in the regulation of mammalian host's physiological processes, it was hypothesized that ticks might encode serpins to evade host defense, and that blocking their function *via* immunization would compromise the tick's ability to feed (Mulenga et al., 2001; Mulenga et al., 2003). Indeed, a limited number of studies have reported mortality and reduced feeding efficiency in *Ixodes ricinus* Say (Prevot et al., 2007), *Haemaphysalis longicornis* Neuman (Sugino et al., 2003; Imamura et al., 2005) and *Rhipicephalus appendiculatus* Neuman (Imamura et al., 2005) ticks that fed on recombinant-serpin-immunized hosts. As part of our long-term

study to explore the role of serpins in tick physiology, the objective of the current study was to identify and characterize *Amblyomma americanum* L. encoded serpins that are expressed early on, during the preparatory and slow feeding phase. We have used serpin generic primers and a previously published PCR approach (Mulenga et al., 2003) to identify and characterize 17 complete and two partial sequences of *A. americanum* serpin variants, here named Lospins, an acronym representing the 'Lone star tick serpin'. We have identified a cluster of ten highly related Lospins that are conserved in several tick species.

The lone star tick *A. americanum* is among the most important and commonly encountered pests of humans and livestock in the Southern USA (Kollars et al., 2000). With its expanding range (Keirans and Lacombe, 1998; Merten and Durden, 2000) across the USA and its role as a vector of important human pathogens, including *Elicchia chaffeensis*, *E. ewingii* and *Borrelia lonestari*, *A. americanum*, long considered as a nuisance tick in terms of public health, has now been recognized as a major vector of human disease agents in the USA (Childs and Paddock, 2003). The discovery of Lospins in this study provides a framework for future studies to uncover the role of serpins in the physiology of lone star tick and other ticks.

## Materials and methods

### *Tick dissections and total RNA isolation*

Ticks used in this study were obtained from a colony of *Amblyomma americanum* L. ticks that are maintained in the laboratory of Dr Pete Teel, in our department. To feed ticks, females were placed in cells on the back of a calf in the presence of male ticks that had pre-attached for 3 days. This arrangement allowed female ticks to commence feeding within 24 h of being put on the animal. Tick dissections were routinely done as previously published (Mulenga et al., 2003). Briefly, ticks fed on cattle for 5 days were washed in 70% ethanol, held on glass slides with a pair of soft tissue forceps and their edges trimmed off using a sharp, sterile razor blade. Under a dissection microscope, the dorsal cuticle flap was lifted and salivary glands (SG), midgut (MG) and ovary (OV) were teased out from the ticks using an 18-gauge needle and a soft tissue forceps. All dissected tissues, including the carcass (CA) representing the tick remnant after removal of SG, MG and OV, were stored in RNA later (Ambion, Austin, TX, USA) at  $-80^{\circ}\text{C}$  until used for RNA extraction.

Extraction of total RNA from whole ticks and dissected tick organs was done using the Trizol (Invitrogen, Carlsbad, CA, USA) reagent as previously described (Mulenga et al., 2003). Briefly, within the first hour of being detached from the host, whole ticks that were partially fed for 24 h (25 ticks), 96 h (10 ticks) and 120 h (10 ticks) were rinsed in 70% ethanol, pulverized in liquid nitrogen, and transferred to the Trizol reagent for RNA extraction. Similarly, tick organs dissected from 20 ticks that were partially fed for 5 days, were rinsed in DPEC water to remove the storage solution and then transferred to the Trizol reagent for RNA extraction. Tissue lysis was accomplished either by repeated pipetting (SG, MG, OV) or homogenization (CA) using a Sonic Dismembrator Model 100 (Fisher Scientific, Pittsburgh, PA, USA). Extracted total RNA was reconstituted in RNase-free water and stored at  $-80^{\circ}\text{C}$  until used.

### *Discovery of lone star tick serpins 'Lospins'*

Cloning was done using generic serpin primers (GSPs) (5'-CATCCTGAACGCTGTCTACTTCAAGGG-3', 5'-CGCGTC-GGCCCTGGAGATACCGTAC-3', 5'-CGTCGACGTTCTC-GACCTGCCTAC-3') in combination with the SMART rapid amplification of cDNA ends kit (RACE; Clontech, San Jose, CA, USA) as published (Mulenga et al., 2003). GSPs were designed based on conserved nucleic acid sequences that were revealed by a multiple sequence alignment (not shown) of annotated tick serpin cDNA sequences from *R. appendiculatus* (Mulenga et al., 2003) *H. longicornis* (Sugino et al., 2003) *Ixodes ricinus* Leach (Prevot et al., 2006), *I. scapularis* (Ribeiro et al., 2006) and *Boophilus microplus* Canestrini (AAP75707). In order to amplify tick serpin genes that are expressed early during the tick feeding cycle, the cDNA template primed by an adapter-linked oligodT primer (Clontech) was synthesized from 5 µg of total RNA extracted from a mixture of ticks that were partially fed for 24 h, 96 h and 120 h.

In the first round of sequencing, 56 PCR fragments cloned in pCR4-TOPO plasmid (Invitrogen) were sequenced and it was established that ~80% of cDNAs encoded a serpin-like polypeptide, as revealed by BLASTX homology search. In the second round of sequencing, 288 insert positive clones were submitted to SequenceWright (Fisher Scientific, Houston, TX, USA) for high throughput sequencing and contig assembly. Following contig assembly and singleton identification, gene-specific PCR primers were designed and used with the 5' and 3' RACE, to clone full-length cDNAs.

### *DNA sequence analyses*

DNA sequences were routinely analyzed using the Vector NTI software packages (Invitrogen, free academic license). For comparison with known serpins and provisional identification, cDNA sequences were scanned against known protein entries in GenBank using the BLASTX and BLASTP homology search program. Additionally, deduced amino sequences were submitted to the ExPASy Proteomics Server (<http://ca.expasy.org/>) for prediction of signal peptides, amino acid motifs and patterns.

### *Structure-based alignment, comparative modeling and calculation of electrostatic surface potential*

In order to predict secondary structures, structure-based alignment was performed between deduced Lospin amino acid residues and the native monomer of antithrombin (1AZX, chain I) using Expresso (Armougom et al., 2006). The 1AZXi template was retrieved from the protein data bank (PDB) as a molecular template based on its 31% and 51% amino acid sequence identity and similarity to Lospins, respectively, and the fact that its reactive centre loop (RCL) was resolved. The RCL, the region of the serpin molecule that is responsible for interaction with target proteinases, forms an extended, exposed conformation above the body of the serpin scaffold (Huntington, 2006; Gettins, 2002). Sequence alignments were subsequently used as input in the MODELLER version 9v1 (Sali and Blundell, 1993) to predict comparative models. The models obtained were evaluated using Verify3D (Luthy et al., 1992) and PROCHECK (Morris et al., 1992). The electrostatic potential of antithrombin (1AZXi, positive control, template),

Table 1. Gene specific primers used in semi-quantitative RT-PCR

Gene ID	Forward primer	Reverse primer
Lospin 1	<sup>S1/2/3</sup> TTGTGCTCTTCACCGCAGCCGTGATG	CCACGGTTCCTTCTTCGTTTACTTC
Lospin 2	<sup>S1/2/3</sup> TTGTGCTCTTCACCGCAGCCGTGATG	CAAGGCTGATCCCCGATAAGTCTGC
Lospin 3	<sup>S1/2/3</sup> TTGTGCTCTTCACCGCAGCCGTGATG	CTGGAGGGGTGGCAAACGCGCCTTC
Lospin 4	ATGTTCTCCAAGTTGGTATTTCTGGCG	GCTCGCATGGCGGGCACTAGGCCG
Lospin 5	<sup>S15/16</sup> CAGGGACGGTCTGTCTAGCG	CAG CCG ATT GTC TAC CGT GGC
Lospin 6	<sup>S17/6</sup> CATGGTCGTCTTGCTTCCAGAC	GTCATTCTGGAAATAGAAGAGGAG
Lospin 7	GGATCCATGTCGGAAGCCATGGCGG	CATCCGTTACTGACCATCCCACTC
Lospin 8	<sup>S9/8</sup> GGATCCCAAGAGGAGCAAAAAGGTGG	GAGTGTAGATGATGACACCTGTGACG
Lospin 9	<sup>S9/8</sup> GGATCCCAAGAGGAGCAAAAAGGTGG	GCTCTGAGTGTACATGATGACACCG
Lospin 11	GAAGACGTCGACAGCAAGCGAG	CTTTGCTTACCTTACAATTTAACTTTATGC
Lospin 13	<sup>S13/14</sup> CCCAAGTTCGACATGAGCCTTC	CGGAGGACGATGGCTCCCATCTC
Lospin 14	<sup>S13/14</sup> CCCAAGTTCGACATGAGCCTTC	GAAATAGAAGAGGAGGTTTCGTG
Lospin 15	<sup>S15/16</sup> CAGGGACGGTCTGTCTAGCG	CATATTAGCCGATTGTCTGGCTTC
Lospin 16	<sup>S15/16</sup> CAGGGACGGTCTGTCTAGCG	GGGCAAGCAGTGGTATCAATTG
Lospin 17	<sup>S17/6</sup> CATGGTCGTCTTGCTTCCAGAC	CGATTGTCTACCGTGGCAGAGC

<sup>S1/2/3</sup>Common forward primer used for Lospins 1–3.  
<sup>S15/16</sup>Common forward primer used for Lospins 5, 15 and 16.  
<sup>S9/8</sup>Common forward primer used for Lospins 8 and 9.  
<sup>S17/6</sup>Common forward primer used for Lospins 17 and 6.

PAI-2 (1BY7, negative control) and Lospin models were calculated by the Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al., 2001). Protonation states were assigned using the parameters for solvation energy (PARSE) force field (Sitkoff et al., 1994) for each structure by PDB2PQR (Dolinsky et al., 2004). Execution of APBS and visualization of resulting electrostatic potentials were performed by PyMol 0.99rev10 (DeLano, 2002) at ±5 kT/e of positive and negative contour fields.

Phylogeny tree construction and similarity comparisons

The phylogeny tree out rooted from the serpin superfamily archetype, human α-1 antitrypsin (AAB59495), was constructed from the dataset of 15 tick serpin polypeptide sequences downloaded from GenBank (accession numbers shown in Fig. 5) and 17 Lospin variants from this study using the neighbor joining method. Specifications were set for bootstrap values at 1000 replications, gaps proportionately distributed and correction for distance set to a Poisson distribution. Amino acid sequence identities among Lospins and other tick serpin polypeptides were determined by pairwise alignment using the Vector NTI software package.

Expression analysis by semi-quantitative RT-PCR

In order to determine spatial patterns of expression for cloned Lospin genes, DNase treated SG, OV, MG and CA total RNA was subjected to two-step RT-PCR using gene specific primers (GSPs, Table 1). OligodT primed first strand cDNA templates were synthesized from ~5 µg DNase treated total RNA using the first strand synthesis kit (Invitrogen). DNase treatment of total RNA was accomplished by a 45 min incubation at 37°C with 1 U RQ1 DNase (Promega, Madison, WI, USA) per 10 µg of RNA, followed by a standard Trizol reagent extraction. A 1 µl aliquot of the first strand cDNA template was used in a PCR reaction with GSPs that were designed based on variable domains of candidate Lospins. A 15 µl aliquot of the PCR

product was electrophoresed on a 2% agarose gel containing 1 µg ethidium bromide. To determine transcript abundance, densitograms of amplified PCR bands were determined using the web based ImageJ image analyzer software (<http://rsb.info.nih.gov/ij/>). To correct for differences due to variations between template concentrations, densities of detected PCR bands were normalized according to the following formula:  $Y=V+V(H-X)/X$ , where  $Y$ =normalized mRNA density,  $V$ =observed Lospin PCR band density in individual tissues (MG, SG and OV),  $H$ =highest tick 16S rRNA PCR band density among tested tissues (carcass in this case, CA),  $X$ =tissue (MG, SG and OV) tick 16S rRNA PCR band density.

Results

Discovery and provisional identification

A previously published PCR based approach (Mulenga et al., 2003) and high throughput sequencing were successfully used to clone 234 serpin cDNA fragments that segregated into 17 contigs and two singletons for a total of 19 partial serpin-like cDNA sequences, here named Lospins (L), an acronym representing the Lone Star tick serpin. The 17 full-length serpin cDNAs have been deposited in GenBank (accession numbers in ascending order, from L1 to L17, are EU072726, EU072727, EU072728, EU072729, EU072730, EU072731, EU072732, EU072733, EU072734, EU072735, EU072736, EU072737, EU072738, EU072739, EU072740, EU072741, EU072742). Consistent with the size of a typical serpin (Gettins, 2002), all deduced Lospin proteins range between 370 and 400 amino acid residues (not shown). BLASTX homology scanning against known protein entries in GenBank was used for routine provisional identification and revealed that all deduced polypeptides in this study showed high similarity exclusively to annotated serpins from other ticks (results not shown). However, when serpin sequences from other ticks were excluded from consideration, best matches included the leukocyte/monocyte elastase inhibitor, neuroserpin, squamous

$\alpha_1$ -ATT	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	E%C
F <sup>33</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	79
N <sup>49</sup>	-	-	-	-	-	-	-	-	-	-	P	-	S	S	-	-	-	87
S <sup>53</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	93
P <sup>54</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	90
S <sup>56</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	72
L <sup>61</sup>	-	-	-	F	F	F	-	-	-	-	F	F	F	F	F	F	F	75
G <sup>67</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	80
T <sup>72</sup>	-	-	-	S	S	S	-	-	-	-	S	S	S	S	S	S	S	87
L <sup>80</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	75
F <sup>130</sup>	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	75
F <sup>147</sup>	-	-	-	-	-	-	-	L	L	L	-	-	-	-	-	-	-	84
I <sup>157</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	83
N <sup>158</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	94
V <sup>161</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	78
T <sup>165</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	89
G <sup>167</sup>	-	-	-	-	-	-	S	-	-	-	-	-	-	-	-	-	-	75
I <sup>169</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	84
T <sup>180</sup>	-	-	-	M	M	V	-	A	A	A	V	M	M	M	M	M	V	75
L <sup>184</sup>	-	-	-	I	I	I	-	-	-	-	I	I	I	I	I	I	I	74
N <sup>186</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	85
F <sup>190</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95
K <sup>191</sup>	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	78
G <sup>192</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	74
W <sup>194</sup>	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	94
F <sup>198</sup>	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95
T <sup>203</sup>	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	84
P <sup>208</sup>	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	98
V <sup>218</sup>	-	-	-	K	K	K	-	-	-	-	K	K	K	K	K	K	K	80
M <sup>220</sup>	T	T	T	A	A	A	-	-	-	-	A	A	A	T	A	A	A	84
M <sup>221</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	86
Y <sup>244</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	76
L <sup>254</sup>	-	-	-	-	-	-	-	R	R	R	-	-	-	-	-	-	-	80
P <sup>255</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	93
P <sup>289</sup>	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	96
K <sup>290</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	72
L <sup>299</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	79
L <sup>303</sup>	-	-	-	M	M	M	-	-	-	-	M	M	M	M	M	M	M	90
G <sup>307</sup>	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	Q	-	83
F <sup>312</sup>	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	90
A <sup>316</sup>	-	-	-	P	-	-	-	-	-	-	S	-	-	-	-	V	-	80
L <sup>327</sup>	-	-	-	R	V	T	I	-	-	-	I	V	T	I	T	T	T	72
H <sup>334</sup>	-	P	P	-	-	-	-	-	-	-	-	-	-	-	-	S	-	78
E <sup>342</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	91
G <sup>344</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	89
A <sup>347</sup>	-	-	-	-	-	-	-	-	-	-	P	-	P	P	P	P	-	79
P <sup>369</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	96
F <sup>370</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	97
L <sup>383</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	80
F <sup>384</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	94
G <sup>386</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	89
P <sup>391</sup>	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	95

Fig. 1. Conservation of core serpin superfamily residues (Irving et al., 2000) in deduced Lospin proteins. The serpin superfamily archetype,  $\alpha_1$ -antitrypsin ( $\alpha_1$ -ATT, accession no. AAB59495) was aligned with deduced Lospin (L) proteins using Vector NTI (Invitrogen) and conserved residues were manually aligned. L1–L17=Lospins 1–17. The minus sign (–) denotes conservation of the specific amino acid residues. Replaced residues are indicated. Numbering is based on  $\alpha_1$ -ATT. E%C denotes the expected % conservation at that position (Irving et al., 2000).

carcinoma antigen 2, plasminogen activator inhibitor 2 (PAI-2) and antithrombin (results not shown).

#### Sequence analysis and structural based alignment

Scanning for signal peptides using SignalP (Emanuelsson et al., 2007) revealed that except for L2, L3, L7 and L11, all other deduced proteins possess leader sequences, indicating they are potentially secreted proteins. Signal peptidase cleavage sites are

predicted after position (p) 21 for L1 and p16 for L4–6, L8–10 and L12–17 (not shown). When scanned for amino acid sequence patterns on the ScanProsite (de Castro et al., 2006), all deduced Lospin proteins were predicted to have multiple potential *N*-glycosylation sites (NX [T/S]) (results not shown). Except for L2, L3 and L11, all other Lospin sequences contain the serpin signature motif pattern PS00284 ([LIVMFY]–[G]–[LIVMFYAC]–[DNQ]–[RKHQS]–[PST]–F–[LIVMFY]–



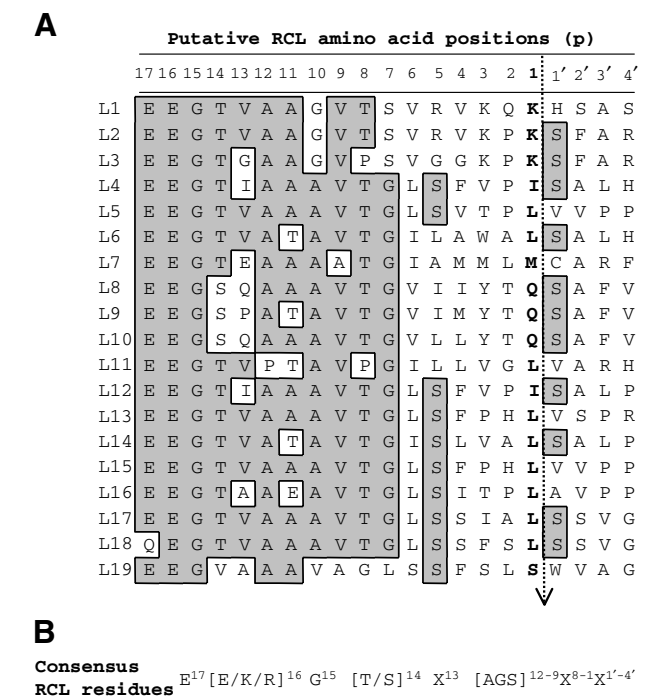


Fig. 2. (A) Predicted Lospin reactive center loops (RCL). Lospin RCLs were determined based on the eight-residue pattern (B) that characterizes inhibitory serpins (Hopkins et al., 1993). The residues in the RCL are numbered according to the standard nomenclature (Schechter and Berger, 1967), where residues on the amino-terminal side of the scissile (p1–p1') are not primed and those on the carboxy-terminal side are primed. Assuming that there are 17 residues between the base of RCL hinge and the scissile bond (Hopkins et al., 1993), bold-faced residues are predicted p1 residues with the broken arrow indicating the position of the scissile bond. L1–L17=Lospins 1–17.

[LIVMFYC]–x–[LIVMFAH]) (results not shown). Four of the 17 Lospins (L1 and L8–10), are predicted to contain the 'TKL' and 'NHL' microbody C-terminal targeting signal pattern PS00342 ([STAGCN]–[RKH]–[LIVMAFY]) at their C-terminal end. When scanned on the 2ZIP-Server (Bornberg-Bauer et al., 1998) (<http://2zip.molgen.mpg.de/index.html>), all Lospins except for L8–10 are predicted to contain leucine residue repeats, [L–(x4)–L–(x4)–L–(x4)–L] that are similar to, but are predicted not fold into, leucine zipper DNA binding patterns (Bornberg-Bauer et al., 1998).

Similarity and identity comparisons between Lospin deduced proteins and human  $\alpha$ -1 antitrypsin (accession no. AAB59495) revealed that 51 core residues occupying strategic buried positions to maintain the overall structure and facilitate the inhibitory mechanism of a serpin molecule (Irving et al., 2000) are 72–98% (38–48/51) conserved in Lospins (Fig. 1). One of the structural features that facilitate the swift function of the serpin molecule is the shutter region (Irving et al., 2000; Hopkins et al., 1993). This region is characterized by conserved amino acid motifs: S<sup>53</sup>–P<sup>54</sup>–X<sup>55</sup>–P<sup>56</sup> (numbering is based on the serpin superfamily archetype; human  $\alpha$ -1 antitrypsin) and I<sup>157</sup>–N<sup>158</sup>–X<sup>159</sup>–X<sup>160</sup>–V<sup>161</sup> (Irving et al., 2000), which are 100% conserved in all Lospin polypeptides (Fig. 1).

On the basis of conservation of the consensus amino acid motif [p17 (E), p16 (E/K/R), p15 (G), p14 (T/S), p12–9 (AGS)] in the hinge region of the RCL that is conventionally used to distinguish between inhibitory and non-inhibitor serpins at the sequence level (Hopkins et al., 1993), all Lospin deduced proteins are putatively inhibitory (Fig. 2). Except for L11, where p12 (A/G/S) is replaced by 'p', L18 where p17 (E), has been replaced by 'Q', and L19, where p9 (A/G/S) is replaced by 'V', all other residues are 100% conserved in the hinge region of putative Lospin RCLs (Fig. 2). The p8 position, which has a high preference for the small threonine side chain (Gettins, 2002), is 100% conserved in all Lospins, except for L3 and L11, where there is a 'P' replacement. Assuming that there are 17 residues between the scissile bond (p1–p1') and the hinge region of the RCL (Hopkins et al., 1993) the predicted p1 residues are 'K', for L1–3, 'I', for L4 and L12, 'L' for L4–6, L11 and L13–18, 'M' for L7, 'Q' for L8–10 and 'S' for L19 (Fig. 2). Overall, when compared to each other at the RCL level, amino acid residue identities ranging from 19–95% were observed (not shown).

Given the conservation of key amino acids that underpin the structure and functionality of serpins, we performed structure-based alignment to gain insight on the putative secondary structures of Lospins. Consistent with the common fold of a typical serpin (Huntington, 2006), structural alignment with the monomer for native antithrombin revealed that each Lospin tertiary structure possess three  $\beta$ -sheets (A–C), eight  $\alpha$ -helices and a reactive center loop (RCL) (Fig. 3).

*Lospins 5 and L13–16 possess basic patches similar to antithrombin (AT) heparin binding site*

Visual inspection of deduced Lospin amino acid sequences revealed two clusters of basic residues (bold), <sup>86</sup>DKSRVLKAYKRL<sup>97</sup> present in L5 and L13–16, and <sup>158</sup>VRDKTRGKI<sup>166</sup>, present in all Lospins (not shown). These patterns showed similarity to glycosaminoglycan (GAG) binding sites; XBnXmB (where X and B=non-basic and basic residues respectively, n=1, 2, or 3 basic residues and m=1 or 2 non-basic residues (Munoz and Linhardt, 2004; Olenina et al., 2005). On three-dimensional models, the <sup>86</sup>DKSRVLKAYKRL<sup>97</sup> and <sup>158</sup>VRDKTRGKI<sup>166</sup> motifs mapped onto  $\alpha$ -helices D and F, respectively (not shown). It was interesting to note that the spatial arrangement for K<sup>87</sup>, K<sup>92</sup> and R<sup>96</sup> in  $\alpha$ -helix D of L5 and L13–16 was comparable to the three residues on  $\alpha$ -helix D of antithrombin, K<sup>114</sup>, K<sup>125</sup> and R<sup>129</sup>, which are important in heparin binding (Olson et al., 2002; Schedin-Weiss et al., 2004; dela Cruz, 2006). To further examine the possibility of the basic residues on  $\alpha$ -helix D of L5 and L13–16 being involved in heparin (GAG) binding activity, we calculated surface electrostatic potentials for L5 and L13–16 models. This analysis revealed that comparative models of L5 and L13–16 possess basic patches (Fig. 4D–F) that are comparable to that of antithrombin (Fig. 4A). Lopin 7 (Fig. 4C), which possess basic residues on its  $\alpha$ -helix F, but lacks the <sup>86</sup>DKSRVLKAYKRL<sup>97</sup> on its  $\alpha$ -helix D, has a much smaller basic patch. The plasminogen activator inhibitor-2, which does not possess basic residues on  $\alpha$ -helix D was used as a negative control and does not possess a basic patch (Fig. 4B).

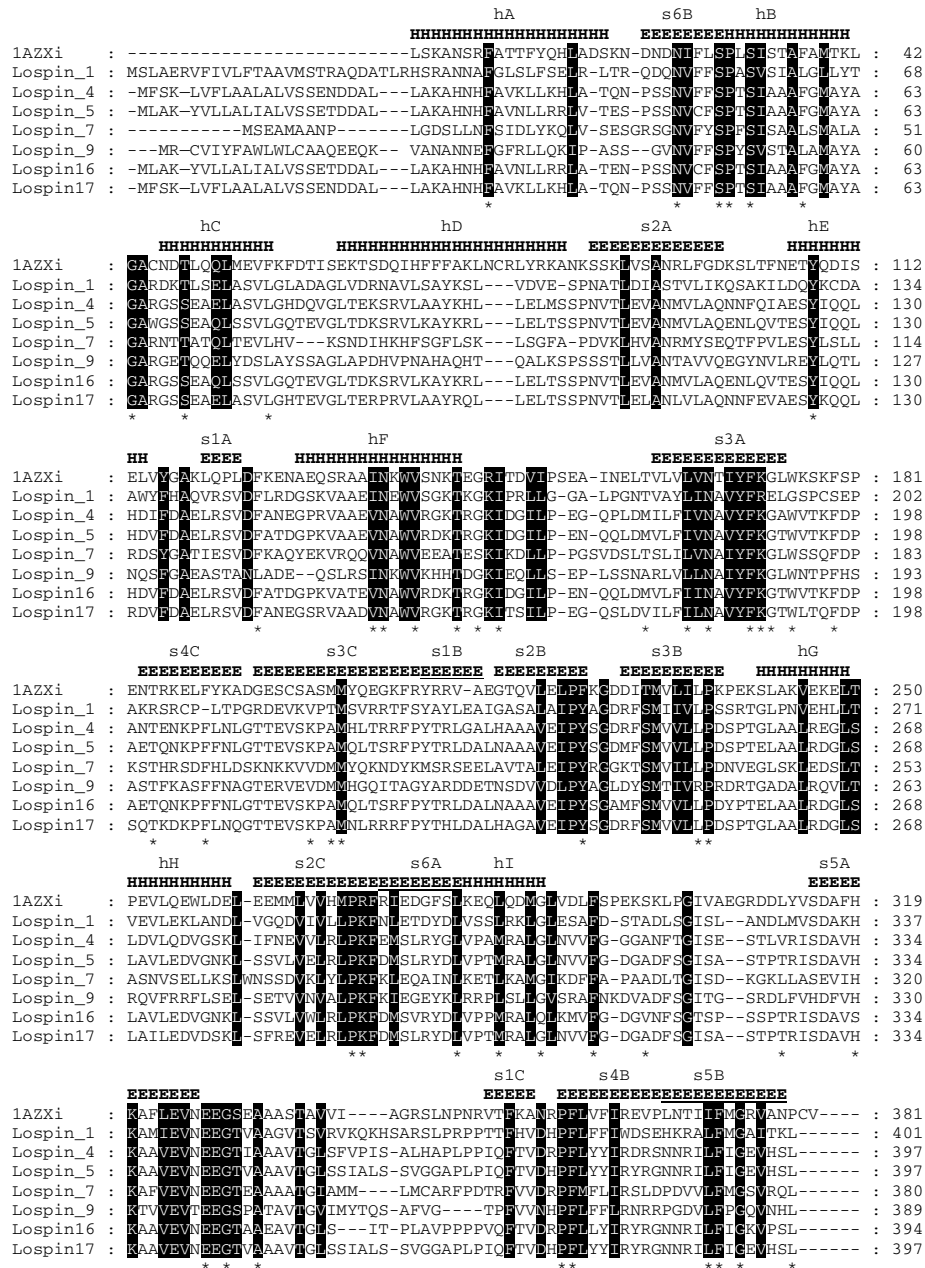


Fig. 3. Structure-based sequence alignment of Lospins with native antithrombin (1AZX, chain I). Structure-based pairwise alignments between Lospins deduced and 1AZX amino acid sequences retrieved from the protein data bank (PDB), were done using Expresso (Armougom et al., 2006). Due to limitations on space, representative alignments only are shown here. Secondary structures were assigned based on 1AZXi (PDB). 'H',  $\alpha$ -helix; 'E', beta-strand. Helices are labeled from 'hA' to 'hI',  $\beta$ -strands that constitute  $\beta$ -sheet A are labeled as 'sA', 'sB' for  $\beta$ -sheet B and 'sC' for  $\beta$ -sheet C. Highly conserved residues that correspond to the 51 core residues shown in Fig. 1 are indicated by an asterisk (\*). Please note that s3c and s1B, s2C and s6A as well as s4B and s5B are merged and continuous, to denote boundaries; s1B, s6A and s5B are underlined.

#### Phylogeny tree and sequence similarity comparison

To determine the relationship among tick serpins, 17 Lospin polypeptides and 15 serpins from other ticks were subjected to phylogeny analysis using the neighbor joining method. From the  $\alpha_1$ -antitrypsin outlier, the aligned sequences segregated into five major groups (A–E) that are supported by bootstrap values of 76% for group A, 100% for groups B, D and E as well as 99% for group C (Fig. 5). In group A, L7 is distantly related from other Lospin proteins, segregated together with the *I. ricinus*

immunosuppressor protein (Iris) (Prevot et al., 2006), *R. appendiculatus* serpin (Ras) 1 and 2 (Mulenga et al., 2003) and *B. microplus* serpins (Bmserpin) 1 (TC8000), 3 (CV44398) and 5 (TC10590). In group B, L8–10 segregated together with Ras-4 (Mulenga et al., 2003) while serpin sequences from *I. ricinus* [serpins 1 (ABI94055), 2 (ABI94056) and 4 (ABI94057)] and *I. scapularis* (AAV80788) in group C are not closely related with any of the Lospin sequences. The majority of the Lospin polypeptides, L4–6 and L11–17, segregated together with Ras-

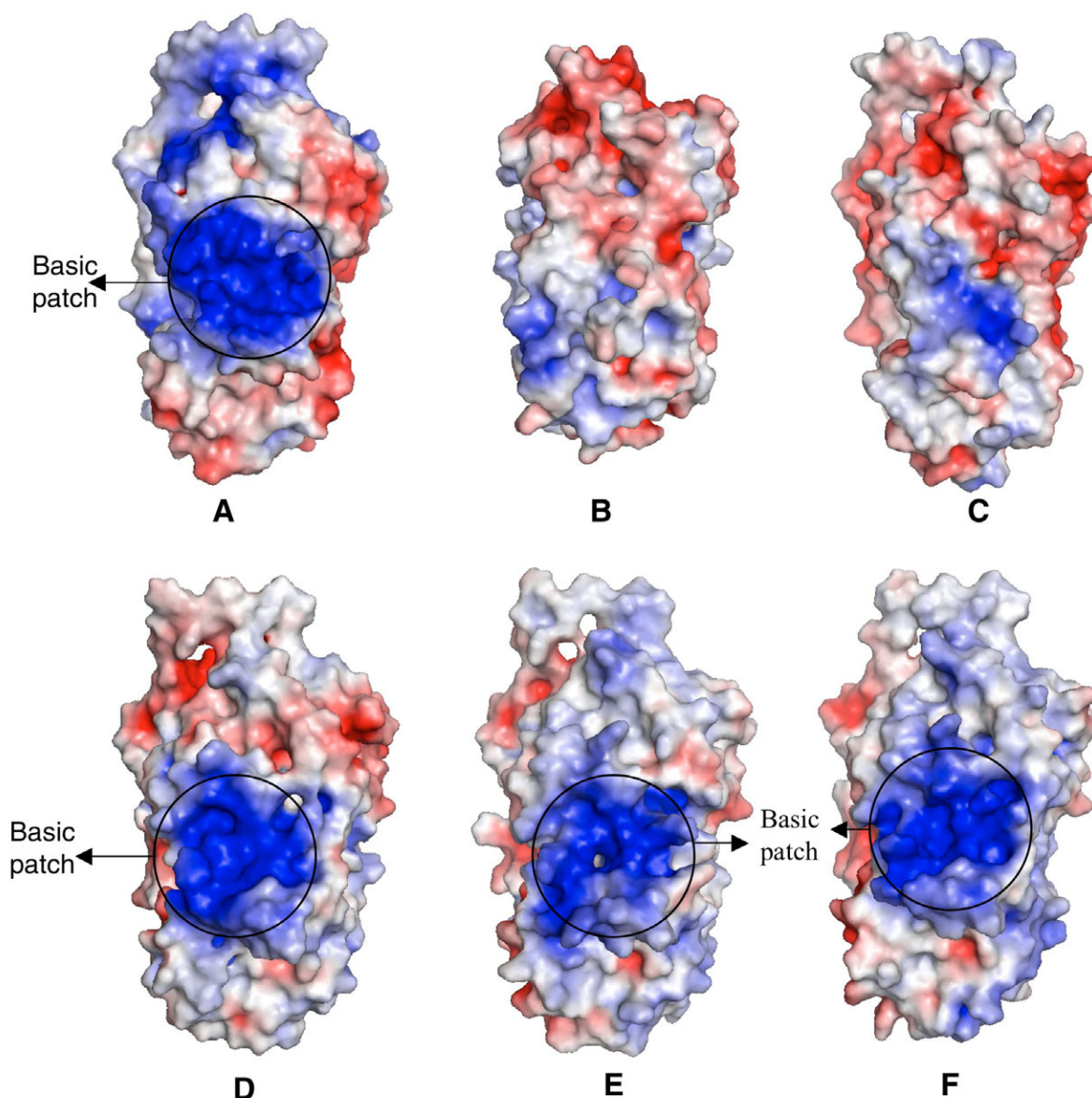


Fig. 4. Comparative modeling and calculation of surface electrostatic potential for L5 and L13–16. Structure-based alignments from Fig. 3 were used as input in the MODELLER version 9v1 (Sali and Blundell, 1993) to develop Lospin models, which were subsequently verified as described in Materials and methods. Electrostatic potential of the template, 1AXZi ('A', native antithrombin), 1BY7 ('B', plasminogen activator inhibitor-2, negative control) and Lospin models were calculated as described in Materials and methods and visualized using PyMol 0.99rev10 [(DeLano, 2002) [www.pymol.org](http://www.pymol.org)] at  $\pm 5\text{kt/e}$  of positive and negative contour fields. Structures D, E and F denote Lospins 5, 13 and 16 models. Basic patches are indicated and marked by a solid circle. C=Lospin 7 (control, has no basic residues on the  $\alpha$ -helix D). Structures are shown back to front.

3 (Mulenga et al., 2003), Bmserpins 2 (TC7417), 4 (CV450507) and 6 (AAP75707), as well as *H. longicornis* (Hl) serpin (BAD11156) in group D, while L1–3, in group E, did not cluster with serpin sequences from other ticks (Fig. 4). Numbering of Bmserpins used in this study is arbitrary. Except for Bmserpins 6, which is annotated in GenBank, the rest of the *B. microplus* serpins used in this study were obtained from the EST database available at 'www.tigr.org'. Percent identity analyses at amino acid level revealed that among group 'A' members, L7, which shows ~33% identity to other Lospins, is 66% and 65% identical to *R. appendiculatus* serpin (Ras) 1 and 2 (AYO35779 and AYO3535780), respectively, 63% to *I. ricinus* blood meal induced immunosuppressor (Iris, CAB55818) and 44–68% to *B.*

*microplus* serpins 1, 3, 5 (TC8000, CV44398, TC10590, respectively) (not shown). While amino acid identity levels of between 93–96% were observed among group B Lospins, L8–10 are 23–43% identical to Ras-4 (Mulenga et al., 2003). Alignment of group D members revealed that these serpins were highly conserved across several tick species with similarity levels of between 74–96% being observed among Lospins, 56–70%, 54–62%, 56–72% and, 54–69% identity being observed when Lospins were compared to Ras-3, Hlserpin, Bmserpins 2, 4 and 6, respectively. Among group E members, L1–3 identity levels of between 83–86% were observed (not shown).

Examination of pairwise alignments among some group D members revealed interesting amino acid residue similarity and



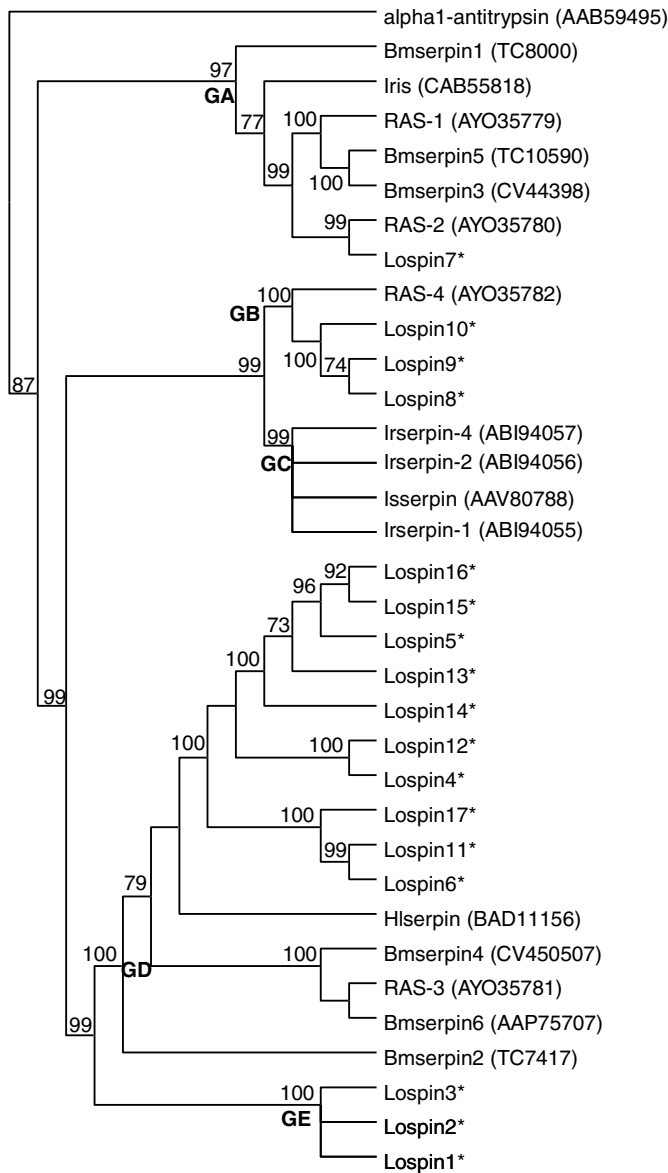


Fig. 5. Neighbor-joining guide phylogeny tree showing the relationship between 17 deduced Lospin polypeptides and 15 serpin sequences of other ticks. Deduced Lospin amino acid sequences were aligned with *R. appendiculatus* (Ras-1–4), *H. longicornis* (Hl) serpin, *I. ricinus* (Ir) serpin 1, 2, 4 and Iris=immunosuppressive *Ixodes* serpin, *I. scapularis* (Is) serpin and *B. microplus* (Bm) serpin 1–6. Note that, except for Bmserpin-5, which is annotated in GenBank, the other Bmserpins were obtained as ESTs from the TIGR database and translated in this study. Branch labels GA–GE, represent major groups of serpin sequences branching off from the outlier, the serpin superfamily archetype,  $\alpha_1$ -antitrypsin. \*Accession numbers for Lospin sequences provided in the Results.

identity patterns where differences between polypeptides are confined to one segment of the sequence. The L6 and L17 alignment revealed differences confined to the first 64 amino-terminal and the last 134 carboxy-terminal (CT) amino acids, with the central domains being identical (not shown). Similarly for L15 and L16, the first 269 amino-terminal residues are identical, with differences confined to the last 125 CT amino

acids (not shown). Patterns comparable to L15 and L16 were observed when any two of the following sequences, L5, L13, L14, L15 and L16, were aligned (not shown).

#### *Lospins are ubiquitously expressed*

To get an insight into tissue distribution profiles of candidate Lospins, gene specific primers based on variable regions of each Lospin cDNA sequences (Table 1) were used to investigate Lospin mRNA expression patterns in SG, MG, OV and CA, dissected from 5-day fed *A. americanum* female ticks. Except for L8, L9 and L17, whose PCR products were not detectable in the OV, and L16, which was not detectable in the CA, the other tested genes are ubiquitously expressed (Fig. 6A). It is interesting to note that for the most part, our RT-PCR expression analysis results were consistent with sequence clustering in the phylogeny tree in Fig. 4. Based on normalized PCR band densities, L1–3, which cluster together in the phylogeny tree (Fig. 4) and show up to 86% amino acid identity, are ~55–70% predominantly expressed in the MG followed by SG (~15–20%), CA (~5–15%), and least expressed in the OV. Similarly, L8 and L9, which also segregated together on the phylogeny tree (Fig. 5) and are 93% identical at the amino acid sequence level (not shown), are ~60–70% predominantly expressed in the MG followed by ~28% expression in the CA, ~3–15% in the SG and no expression in the OV. Comparable to L1–3, L7 is ~55% highly expressed in the MG, ~25% in the CA, ~15% in the SG and least expressed in the OV. Among group D (Fig. 4) tested members, L5, L11 and L16 display superior expression in the OV by ~38%, 50% and 90%, followed by ~25%, 20% and 10% expression in the MG, respectively. Additionally, while L16 expression in the CA and SG was below 1%, L5 and L11, respectively, show ~10% and 15% expression in the CA and 2% and 8% in the SG. Among the other group D tested genes, L17, which is not expressed in the OV is expressed to equivalent levels in the CA and SG, respectively, by ~40% and ~20% in the MG, while L4, L6, L13 and L14, respectively, are ~30%, 38%, 28% and 50% expressed in the CA, ~28%, 32%, 20% and 25% in the SG, ~30%, 20%, 22% and 20% in MG. Additionally, expression in OV is ~10% for L4 and L6, ~18% for L13 and <2% for L14.

#### Discussion

Serpin encoding cDNAs have recently been cloned from several ticks including *Boophilus microplus* (www.tigr.org), *Ixodes scapularis* (Ribeiro et al., 2006), *I. ricinus* (Prevot et al., 2006), *Amblyomma variegatum* (Nene et al., 2002b), *Rhipicephalus appendiculatus* (Mulenga et al., 2003) and *Haemaphysalis longicornis* (Sugino et al., 2003; Imamura et al., 2005). In this study we report on identification, comparative bioinformatics and mRNA expression analyses of 17 full-length and two partial *A. americanum* serpin variants here named 'Lospins'. The Lospin sequences reported were cloned from ticks that had fed for 24 h, 96 h and 120 h. These stages of tick feeding coincide with the preparatory and the slow feeding phases of the tick feeding process, during which the tick establishes its feeding lesion and begins to transmit disease pathogens (Sonenshine, 1993). We are interested in this stage of the tick feeding process because it precedes most of the damage caused by tick feeding activity (Sonenshine, 1993). Our thinking is that, if we block



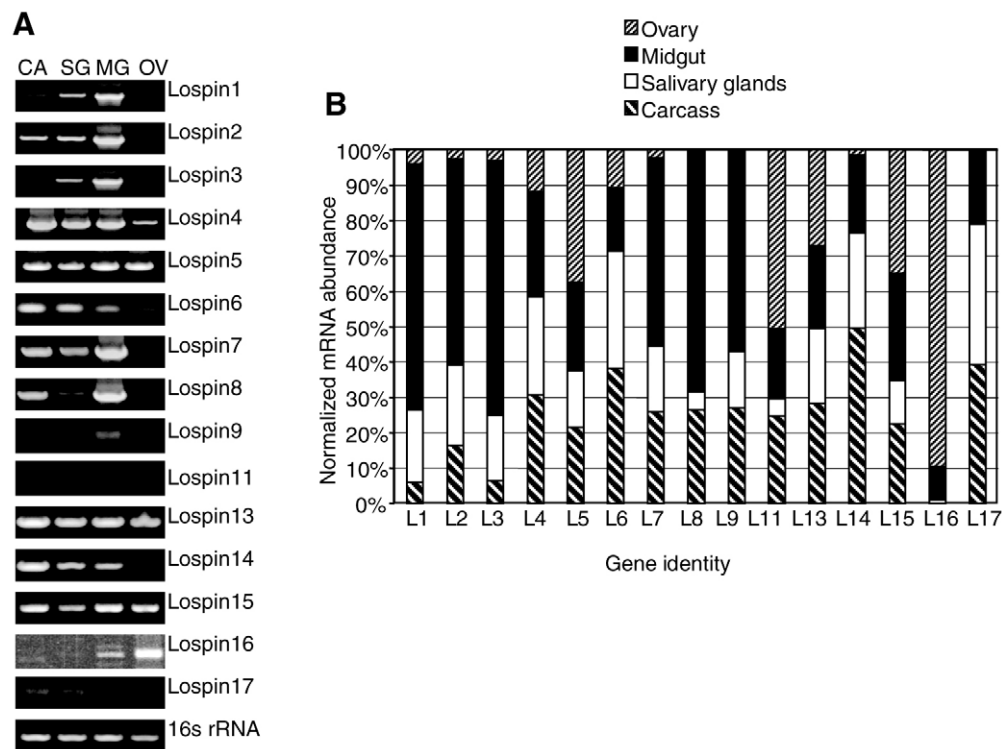


Fig. 6. (A) Transcription profiles and (B) normalized PCR band densities representing relative mRNA abundance of 15 selected Lospins. Total RNA extracted from salivary glands (SG), midgut (MG), ovary (OV) and carcass (CA, remnant of the tick following removal of SG, OV and MG) dissected from 5-day fed *A. americanum* ticks were subjected to semi-quantitative RT-PCR using gene specific primers shown in Table 1. The 16 s rRNA PCR fragment amplified from *A. americanum* 16 s rRNA primers (Table 1) was used as the endogenous control. PCR band densities were determined using ImageJ software (<http://rsb.info.nih.gov/ij>). Determined densities were normalized using the following formula:  $Y = V + V(H - X)/X$ , where  $Y$  = normalized mRNA density,  $V$  = observed Lospin PCR band density in individual tissues (MG, SG and OV),  $H$  = highest 16 s rRNA PCR band density among tested tissues (carcass in this case, CA),  $X$  = tissue (MG, SG and OV) 16 s rRNA PCR band density. (B) Normalized band densities were plotted as percent tissue distributions.

serpin function early enough in the tick feeding cycle, we will not only interfere with the tick's ability to start feeding but also possibly prevent pathogen transmission. The potential limitation of using generic primers to clone multi-member gene families such as serpin is the possibility of a bias against identification of poorly expressed genes. While we are unable to determine whether or not there was a bias against cloning of poorly expressed Lospins, this possibility was minimized by our approach to sequence the entire 344-insert positive clones that were isolated from ligations of our PCR products.

Adoption of the consensus secondary structures of a typical serpin molecule (Gettins, 2002; Huntington, 2006) and the high conservation of the core amino acid residues (Irving et al., 2000) that underpin structure and functionality of serpins strongly suggest that Lospins are functional members of the serpin superfamily. Though originally identified as inhibitors of serine proteinases, cross-class members that can inhibit cysteine proteinases (Pak et al., 2004) and others with no inhibitor functions (Askew et al., 2007) have also been identified. While we are unable to specify the classes of their target proteinases as cysteine or serine, almost all deduced Lospin proteins are predicted to have putative inhibitory functions, as determined by consensus amino acid residues in the hinge regions of their putative RCLs (Hopkins et al., 1993). While confirmatory experimentation is awaited, it is important to point out here that

possession of proline residues at the critical p8 and p12 positions of L3 and L11 deduced RCLs may suggest that these Lospins have no inhibitor functions. In a previous study, point mutations of p12, alanine, p10, serine and p8 threonine to proline resulted in loss of inhibitory activity by plasminogen activator inhibitor-1 (Audenaert et al., 1994). In another study, mutation of the glycine residue at the p10 position to proline converted  $\alpha_1$ -antitrypsin from an inhibitor to a substrate (Hopkins et al., 1993). Deduced RCLs in this study were predicted based on the 17-residue rule (Hopkins et al., 1993; Irving et al., 2000). Given that some characterized serpins such as  $\alpha_2$ -antiplasmin (Gettins, 2002) or serpin1k from *Manduca sexta* (Li et al., 1999) utilize RCLs that are shorter or longer than the conventional 17 residues, we are interpreting our predicted scissile bonds with caution. Consistent with the fact that almost all known serpins are glycosylated (Whisstock et al., 2005; Law et al., 2006; Silverman et al., 2001; Robertson et al., 2006), our bioinformatics analyses data demonstrated that all deduced Lospin sequences possess potential *N*-glycosylation sites. From the perspective of finding antigens for anti-tick vaccine development, it was encouraging to note that, except for L2, L3, L7 and L11, which do not possess leader sequences, the majority of Lospins are predicted to be extracellular. The significance of this finding is that the majority of Lospins represent potential target antigens for anti-tick vaccine

development, in that they will be accessible to host immune response factors. In humans, intracellular serpins, which are classified as clade B serpins (ov-serpins), have a higher frequency of presence of oxidation-sensitive residues such as methionine and cysteine in RCL, which are not normally exposed to highly oxidative conditions extracellularly (Silverman et al., 2004). It is interesting to note that, L7, one of four putative intracellular Lospins, possesses three methionines at p1, p3 and p4 and a cysteine at the p1' position.

Whether or not the putative GAG binding sites in L5 and L13–16 as well as the microbody C-terminal targeting signal in L1, L8 and L9, are functional, awaits experimentation. We are, however, encouraged by the fact that most known proteins that possess GAG binding motifs are involved in regulation of several important physiological pathways, which if disrupted, could severely compromise the tick's ability to feed. For instance, antithrombin, heparin cofactor II, chemokines and selectin, whose functions are regulated by GAG binding, are involved in mediation of essential biological processes such as blood coagulation, inflammatory response, immune cell migration, tumor cell metastasis and smooth muscle cell proliferation (Munoz and Linhardt, 2004). Similarly in invertebrates, GAG binding proteins were associated with immunity (Kamimura et al., 2006; Robertson et al., 2006) and development (Tollefsen, 2007) in *Drosophila*. Microbodies also known as peroxisomes in vertebrates, glyoxysomes, glycosomes and hydrogenosomes, depending on their chemical composition, are small electron-dense membrane-bound organelles that perform a variety of metabolic functions, including the  $\beta$ -oxidation of fatty acids and the biosynthesis of cholesterol and bile acids in eukaryotes (Gould et al., 1990; Gatto et al., 2000). Occurrence of diseases collectively referred to as 'peroxisome biogenesis disorders' in the case of failure to properly assemble peroxisomes (Gould et al., 1990; Gatto et al., 2000) goes to attest to the importance of these microbodies. Clearly our future goal should be to uncover physiological processes regulated by Lospins.

Similarity and identity patterns among group D Lospins, where differences were confined to one region of the sequence, appear to be consistent with features that characterize alternately spliced genes (AS) (Talavera et al., 2007). The possibility of AS as a source of diversity among Lospins is not unique to ticks in that this phenomenon, first observed in *M. sexta* (Jiang et al., 1996), has been observed in serpins from *C. felis* (Brandt et al., 2004), *D. melanogaster* and *C. elegans* (Kruger et al., 2002). Experiments are currently underway to investigate if Lospin diversity could be attributed to AS of mutually spliced exons. Our long-term interest is to identify tick proteins that can be targeted for rational design of new tick control approaches. Thus our sequence analysis data showing that Lospins in groups A and D were conserved in other ticks is encouraging. Given the huge diversity of ticks that can infest animals (de la Fuente and Kocan, 2003), it will be highly desirable to develop new tick control strategies targeting conserved tick proteins such as groups A and D Lospins, in that a single treatment could protect against several tick species.

Although certain Lospins were apparently over-expressed in certain tick organs, the general trend revealed by RT-PCR expression analysis is that the majority of tested Lospins are ubiquitously expressed. Expression of Lospins in multiple tick

organs underscores their importance in regulation of key tick physiological processes. Curiously, all Lospin genes that were highly expressed in the midgut were poorly expressed in the ovary and *vice versa*. It will be interesting to explore the significance of the differences in transcription profiles. It is also important to point out here that the transcription profile data presented here is based on ticks that had fed for 5 days. Whether the transcription profiles will change during the tick feeding cycle was not determined in the current study. While *A. americanum* ESTs encoding serpin fragments were present in GenBank (Hill and Gutierrez, 2000) ([www.genome.ou.edu](http://www.genome.ou.edu)) at the inception of this project, data presented here represents the first reported attempt to characterize annotated *A. americanum* serpins genes. The discovery of Lospins in this study sets the framework for future studies to understand the role of serpins in tick physiology.

We would like to thank Dr Pete Teel for kindly providing ticks used in this study. Funding for this project was provided by start up funds to Albert Mulenga from the Texas A&M College of Agriculture and Life Sciences, Department of Entomology and the Texas Agriculture Experiment Station.

## References

- Abraham, E. G., Pinto, S. B., Ghosh, A., Vanlandingham, D. L., Budd, A., Higgs, S., Kafatos, F. C., Jacobs-Lorena, M. and Michel, K. (2005). An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc. Natl. Acad. Sci. USA* **102**, 16327–16332.
- Armougoum, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. and Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34**, W604–W608.
- Askew, D. J., Cataltepe, S., Kumar, V., Edwards, C., Pace, S. M., Howarth, R. N., Pak, S. C., Askew, Y., Bromme, D., Luke, C. J. et al. (2007). Serpinb11 is a new non-inhibitory intracellular serpin: common single nucleotide polymorphisms in the scaffold impair conformational change. *J. Biol. Chem.* **10.1074/jbc.M703182200**.
- Audenaert, A. M., Knockaert, I., Collen, D. and Declercq, P. J. (1994). Conversion of plasminogen activator inhibitor-1 from inhibitor to substrate by point mutations in the reactive-site loop. *J. Biol. Chem.* **269**, 19559–19564.
- Bornberg-Bauer, E., Rivals, E. and Vingron, M. (1998). Computational approaches to identify leucine zippers. *Nucleic Acids Res.* **26**, 2740–2746.
- Brandt, K. S., Silver, G. M., Becher, A. M., Gaines, P. J., Maddux, J. D., Jarvis, E. E. and Wisniewski, N. (2004). Isolation, characterization, and recombinant expression of multiple serpins from the cat flea, *Ctenocephalides felis*. *Arch. Insect Biochem. Physiol.* **55**, 200–214.
- Carrell, R. and Corra, J. (2004). What can *Drosophila* tell us about serpins, thrombosis and dementia. *BioEssays* **26**, 1–5.
- Childs, J. E. and Paddock, C. D. (2003). The ascendancy of *Amblyomma americanum* as a vector of pathogens affecting humans in the United States. *Annu. Rev. Entomol.* **48**, 307–337.
- de Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365.
- de la Cruz, R. G. C., Jairajpuri, M. A. and Bock, S. C. (2006). Disruption of a tight cluster surrounding tyrosine 131 in the native conformation of antithrombin III activates it for factor Xa inhibition. *J. Biol. Chem.* **281**, 31668–31676.
- de la Fuente, J. and Kocan, K. M. (2003). Advances in the identification and characterization of protective antigens for recombinant vaccines against tick infestations. *Expert Rev. Vaccines* **2**, 583–593.
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific. <http://www.pymol.org>.
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.
- Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP, and related tools. *Nat. Protoc.* **2**, 953–971.

- Gatto, G. J., Jr, Geisbrecht, B. V., Gould, S. J. and Berg, J. M. (2000). Peroxisomal targeting signal-1 recognition by the TPR domains of the human PEX5. *Nature* **7**, 1091-1095.
- Gettins, P. W. G. (2002). Serpin structure, mechanism, and function. *Chem. Rev.* **102**, 4751-4803.
- Gould, S. J., Keller, G. A., Schneider, M., Howell, S. H., Garrard, L. J., Goodman, J. M., Distel, B., Tabk, H. and Subramani, S. (1990). Peroxisomal protein import is conserved between yeast, plants, insects and mammals. *EMBO J.* **9**, 85-90.
- Hill, C. A. and Gutierrez, J. A. (2000). Analysis of the expressed genome of the lone star tick, *Amblyomma americanum* (Acari: Ixodidae) using an expressed sequence tag approach. *Microb. Comp. Genomics* **5**, 89-101.
- Hill, C. A. and Wikel, S. K. (2005). The *Ixodes scapularis* genome project: an opportunity for advancing tick research. *Trends Parasitol.* **21**, 151-153.
- Hopkins, P. C., Carrell, R. W. and Stone, S. R. (1993). Effects of mutations in the hinge region of serpins. *Biochemistry* **32**, 7650-7657.
- Huntington, J. A. (2006). Shape-shifting serpins – advantages of a mobile mechanism. *Trends Biochem. Sci.* **31**, 427-435.
- Imamura, S., da Silva, V. J., Sugino, M., Ohashi, K. and Onuma, M. (2005). A serine proteinase inhibitor (serpin) from *Haemaphysalis longicornis* as an anti-tick vaccine. *Vaccine* **23**, 1301-1311.
- Irving, J. A., Pike, R. N., Lesk, A. M. and Whisstock, J. C. (2000). Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res.* **10**, 1845-1864.
- Jiang, H., Wang, Y., Huang, Y., Mulnix, A. B., Kadel, J., Cole, K. and Kanost, M. R. (1996). Organization of serpin gene-1 from *Manduca sexta*. Evolution of a family of alternate exons encoding the reactive site loop. *J. Biol. Chem.* **271**, 28017-28023.
- Kamimura, K., Koyama, T., Habuchi, H., Ueda, R., Masu, M., Kimata, K. and Nakato, H. (2006). Specific and flexible roles of heparan sulfate modifications in *Drosophila* FGF signaling. *J. Cell Biol.* **174**, 773-778.
- Keirans, J. E. and Lacombe, E. H. (1998). First records of *Amblyomma americanum*, *Ixodes (Ixodes) dentatus*, and *Ixodes (Ceraticoxodes) uriae* (Acari: Ixodidae) from Maine. *J. Parasitol.* **84**, 629-631.
- Kollars, T. M., Jr, Oliver, J. H., Jr, Durden, L. A. and Kollars, P. G. (2000). Host association and seasonal activity of *Amblyomma americanum* (Acari: Ixodidae) in Missouri. *J. Parasitol.* **86**, 1156-1159.
- Kruger, O., Ladewi, J., Koster, K. and Ragg, H. (2002). Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes. *Gene* **293**, 97-105.
- Law, H. P. R., Zhang, Q., McCowan, S., Buckle, A. M., Silverman, G. A., Wong, W., Rosado, C. J., Langendorf, C. G., Pike, R. N., Bird, P. I. et al. (2006). An overview of the serpin superfamily. *Genome Biol.* **7**, 216-226.
- Li, J., Wang, Z., Canagarajah, B., Jiang, H., Kanost, M. and Goldsmith, E. J. (1999). The structure of active serpin 1K from *Manduca sexta*. *Structure* **7**, 103-109.
- Luthy, R., Bowie, J. U. and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-85.
- Merten, H. A. and Durden, L. A. (2000). A state-by-state survey of tick recorded from humans in the United States. *J. Vector Ecol.* **25**, 102-113.
- Michel, K., Budd, A., Pinto, S., Gibson, T. J. and Kafatos, F. C. (2005). *Anopheles gambiae* SRPN2 facilitates midgut invasion by the malaria parasite *Plasmodium berghei*. *EMBO Rep.* **6**, 891-897.
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G. and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345-364.
- Mulenga, A., Sugino, M., Nakajima, M., Sugimoto, C. and Onuma, M. (2001). Tick-encoded serine proteinase inhibitors (serpin); potential target antigens for tick vaccine development. *J. Vet. Med.* **63**, 1063-1069.
- Mulenga, A., Tsuda, A., Onuma, M. and Sugimoto, C. (2003). Four serine proteinase inhibitors (serpin) from the brown ear tick, *Rhipicephalus appendiculatus*; cDNA cloning and preliminary characterization. *Insect Biochem. Mol. Biol.* **33**, 267-276.
- Munoz, E. M. and Linhardt, R. J. (2004). Heparin binding domains in vascular biology. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1549-1557.
- Nappi, A. J., Frey, F. and Carton, Y. (2005). *Drosophila* serpin 27A is a likely target for immune suppression of the blood cell-mediated melanotic encapsulation response. *J. Insect Physiol.* **51**, 197-205.
- Nene, V., Lee, D., Quackenbush, J., Skilton, R., Mwaura, S., Gardner, M. J. and Bishop, R. (2002a). Genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *Theileria parva*. *Insect Biochem. Mol. Biol.* **34**, 1117-1128.
- Nene, V., Lee, D., Quackenbush, J., Skilton, R., Mwaura, S., Gardner, M. J. and Bishop, R. (2002b). AvGI, an index of genes transcribed in the salivary glands of the ixodid tick *Amblyomma variegatum*. *Int. J. Parasitol.* **32**, 1447-1456.
- Olenina, L. V., Kuzmina, T. I., Sobolev, B. N., Kuraeva, T. E., Kolesanova, E. F. and Archakov, A. I. (2005). Identification of glycosaminoglycan-binding sites within hepatitis C virus envelope glycoprotein E2. *J. Viral Hepatitis* **12**, 584-593.
- Olson, S. T., Bjork, I. and Bock, S. C. (2002). Identification of critical molecular interactions mediating heparin activation of antithrombin; implications for the design of heparin anticoagulants. *Trends Cardiovasc. Med.* **12**, 198-205.
- Pak, S. C., Kumar, V., Tsu, C., Luke, C. J., Askew, Y. S., Mills, D. R., Bromme, D. and Silverman, G. A. (2004). SRP-2 is a cross-class inhibitor that participates in postembryonic development of the nematode *Caenorhabditis elegans*: initial characterization of the clade L serpins. *J. Biol. Chem.* **279**, 15448-15459.
- Pelte, N., Robertson, A. S., Zou, Z., Belorgey, D., Dafforn, T. R., Jiang, H., Lomas, D., Reichhart, J. M. and Gubb, D. (2006). Immune challenge induces N-terminal cleavage of the *Drosophila* serpin necrotic. *Insect Biochem. Mol. Biol.* **36**, 37-46.
- Potempa, J., Korzus, E. and Travis, J. (1994). The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* **269**, 15957-15960.
- Prevot, P. P., Adam, B., Zouaoui, B. K., Brossard, M., Lins, L., Cauchie, P., Brasseur, R., Vanhaevebeek, M., Vanhamme, L. and Godfroid, E. (2006). Anti-hemostatic effects of a serpin from saliva of the tick *Ixodes ricinus*. *J. Biol. Chem.* **281**, 26361-26369.
- Prevot, P. P., Couvreur, B., Denis, V., Brossard, M., Vanhamme, L. and Godfroid, E. (2007). Protective immunity against *Ixodes ricinus* induced by a salivary serpin. *Vaccine* **25**, 3284-3292.
- Ribeiro, J. M. C., Francisco, A. C., Franchischetti, I. M. B., Mans, B. J., Mather, T. N., Valenzuela, J. G. and Wikel, S. K. (2006). An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem. Mol. Biol.* **36**, 111-119.
- Robertson, A. S., Belorgey, D., Gubb, D., Dafforn, T. R. and Lomas, D. A. (2006). Inhibitory activity of the *Drosophila melanogaster* serpin, necrotic is dependent on lysine residues in the D-helix\*. *J. Biol. Chem.* **281**, 26437-26443.
- Rushlow, C. (2004). Dorsoventral patterning: a serpin pinned down at last. *Curr. Biol.* **14**, R16-R18.
- Sali, A. and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.
- Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.
- Schedin-Weiss, S., Desai, U. R., Bock, S. C., Olson, S. T. and Bjork, I. (2004). Roles of N-terminal region residues Lys11, Arg13, and Arg24 of antithrombin in heparin recognition and in promotion and stabilization of the heparin-induced conformational change. *Biochemistry* **43**, 675-683.
- Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Caughlin, P. B., Gettins, P. G. W., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S. et al. (2001). The serpins are expanding superfamily of structurally similar but functionally diverse proteins: evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J. Biol. Chem.* **276**, 33293-33296.
- Silverman, G. A., Whisstock, J. C., Askew, D. J., Pak, S. C., Luke, C. J., Cataltepe, S., Irving, J. A. and Bird, P. I. (2004). Human clade B serpins (ov-serpins) belong to a cohort of evolutionary dispersed intracellular proteinase inhibitor clades that protect cells from promiscuous proteolysis. *Cell. Mol. Life Sci.* **61**, 301-325.
- Sitkoff, D., Sharp, K. A. and Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978-1988.
- Sonenshine, D. E. (1993). *Biology of Ticks*. Oxford: Oxford University Press.
- Sugino, M., Imamura, S., Mulenga, A., Nakajima, N., Tsuda, A., Ohashi, K. and Onuma, M. (2003). A serine proteinase inhibitor (serpin) from ixodid tick *Haemaphysalis longicornis*; cloning and preliminary assessment of its suitability as a candidate for a tick vaccine. *Vaccine* **21**, 2844-2851.
- Talavera, D., Vogel, C., Orozco, M., Teichmann, S. A. and de la Cruz, X. (2007). The (In)dependence of alternative splicing and gene duplication. *Plos Comput. Biol.* **3**, 1-14.
- Tollefsen, D. M. (2007). Heparin cofactor II modulates the response to vascular injury. *Arterioscler. Thromb. Vasc. Biol.* **27**, 454-460.
- Whisstock, J. C., Bottomley, S. P., Bird, P. I., Pike, R. N. and Coughlin, P. (2005). Serpins 2005 – fun between the  $\beta$ -sheets, meeting report based upon presentations made at the 4th International Symposium on Serpin Structure, function and biology (Cairns, Australia). *FEBS J.* **272**, 4868-4873.
- Zou, Z. and Jiang, H. (2005). *Manduca sexta* serpin-6 regulates immune serine proteinases PAP-3 and HP8. cDNA cloning, protein expression, inhibition kinetics, and function elucidation. *J. Biol. Chem.* **280**, 14341-14348.